Supporting Information

Ren et al. 10.1073/pnas.1005389107

SI Text

SI Methods. ***Protein expression and purification.*** A DNA fragment encompassing full-length ribonucleic acid export 1 (Rae1) (residues 1–368) was amplified by PCR and cloned into the pFastBac Dual vector (Invitrogen), using SmaI and KpnI restriction sites for expression in Sf9 cells (Invitrogen). Rae1 was expressed with a noncleavable C-terminal hepta-histidine tag. The Gle2-binding sequence (GLEBS) motif of Nup98 (residues 157–213) was amplified by PCR and first cloned into a pGEX-2T vector (GE Healthcare) modified to contain a tobacco etch virus (TEV) protease-cleavable N-terminal GST-tag, using BamHI and EcoRI restriction sites. This construct was then used as a PCR template to generate a GST-TEV-Nup98^{GLEBS} DNA fragment flanked by BglII and EcoRI restriction sites. This fragment was cloned into the pFastBac Dual vector that harbored Rae1 in the second expression site, using BamHI and EcoRI restriction sites for coexpresion of the Rae1•Nup98^{GLEBS} complex in Sf9 cells.

For expression of the Rae1•Nup98^{GLEBS} complex, Sf9 cells were infected with a recombinant baculovirus of an approximate multiplicity of infection of five and grown in suspension for three days in Sf-900 II SFM medium (Invitrogen). The cells were pelleted by centrifugation and resuspended in a buffer containing 50 mM Tris, pH 8.0, 300 mM NaCl, 10 mM imidazole, 2 mM β-mercaptoethanol (β-ME), 1 mM phenylmethylsulfonyl fluoride (PMSF), and complete EDTA-free protease inhibitor mixture tablets (Roche). The cells were lysed with a cell disrupter (Avestin), and the lysate was centrifuged for 90 min at 40,000 × *g*. The lysate was then applied onto an HIS-Select column (Sigma) and eluted via an imidazole gradient. Ni^{2+}-affinity chromatography yielded a stoichiometric Rae1•Nup98^{GLEBS} complex, as His-tagged Rae1 was expressed at a lower level than GST-Nup98^{GLEBS}. The GST-tag of Nup98 was removed by TEV protease digestion, leaving an N-terminal overhang with the sequence GS. The complex was further purified over a HiTrap SP column (GE Healthcare) using a 50–200 mM NaCl gradient in 10 mM Tris, pH 8.0. Rae1•Nup98^{GLEBS}-containing fractions were pooled, concentrated, and loaded onto a Superdex 200 10/300 GL gel filtration column (GE Healthcare) and equilibrated with a buffer containing 10 mM Tris, pH 8.0, 150 mM NaCl, and 0.5 mM Tris[2-carboxyethyl]phosphine (TCEP). Purified Rae1•Nup98^{GLEBS} complex was concentrated to 10 mg/mL and flash-frozen in liquid nitrogen for storage at −80 °C. Rae1 alone was produced in Sf9 cells and purified as the Rae1•Nup98^{GLEBS} complex; however, the TEV protease cleavage step was omitted.

A DNA fragment encoding the wild-type Nup98 GLEBS motif was amplified by PCR and cloned in the pGEX-4T1 vector (GE Healthcare) using BamHI and EcoRI restriction sites. Nup98 GLEBS mutants were generated by QuikChange mutagenesis (Stratagene) and confirmed by DNA sequencing. Nup98 GLEBS fragments and variants were expressed in *Escherichia coli* BL21-CodonPlus(DE3)-RIL cells (Stratagene), and protein expression was induced at OD_{600} of 1.0 with 0.5 mM IPTG at 23 °C for 3 h. Cells were pelleted and resuspended in a buffer containing 50 mM Tris, pH 8.0, 150 mM NaCl, 2 mM β-ME, and 1 mM PMSF and lysed with a cell disrupter (Avestin). GST-Nup98^{GLEBS} proteins were purified using a glutathione sepharose 4 fast flow affinity column (GE Healthcare), using a buffer containing 50 mM Tris, pH 8.0, 150 mM NaCl, and 2 mM β-ME and eluted with a gradient of the same buffer supplemented with 20 mM glutathione. Protein-containing fractions were pooled, concentrated, and loaded onto a Superdex 200 10/300 GL gel filtration column (GE Healthcare) in a buffer containing 10 mM Tris, pH 8.0, 150 mM NaCl, and 0.5 mM TCEP. Proteins were concentrated to 2 mg/mL and flash-frozen in liquid nitrogen for storage at −80 °C. The details of the baculoviral and bacterial expression constructs are listed in Table S2.

Crystallization and structure determination. Crystals of the Rae1•Nup98^{GLEBS} complex were obtained at 20 °C by vapor diffusion in hanging drops using 1 μL of the protein (10 mg/mL) and 1 μL of a reservoir solution consisting of 0.1 M MES, pH 6.0, 25% (vol/vol) PEG 2000 monomethyl ether (MME). Crystals grew in the triclinic space group P1, with four complexes in the asymmetric unit, and they grew to their maximum size of 200 × 100 × 30 μm within a week. For cryoprotection, crystals were briefly soaked in 0.1 M MES, pH 6.0, 25% (vol/vol) PEG 2000 MME, and 22 % (vol/vol) glycerol and flash-frozen in liquid nitrogen. Derivatized crystals were obtained by soaking native crystals in the crystallization solution supplemented with 1 mM osmium tetroxide (OsO_4) for 5 min. X-ray diffraction data were collected at the General Medicine and Cancer Institutes Collaborative Access Team (GM/CA-CAT) beamline 23ID-B at the Advanced Photon Source (APS), Argonne National Laboratory. X-ray intensities were processed using the HKL2000 denzo/scalepack package (1), and the CCP4 program package (2) was used for subsequent calculations. A single-wavelength anomalous dispersion X-ray diffraction dataset of an OsO_4-labeled protein crystal was used to identify the positions of 24 osmium atoms with SHELXD (3). Phases were calculated to 3.0 Å in SHARP (4), followed by density modification in DM (2) with solvent flattening and histogram matching. This procedure yielded an electron density map of excellent quality. A model was built with the programs O (5) and Coot (6) and were refined using CNS (7) and Refmac (8). No electron density was observed for Rae1 residues 1–3, 19–22, and 366–368, and Nup98^{GLEBS} residues 157 and 174–178, and these residues have been omitted from the final model. The stereochemical quality of the model was assessed with PROCHECK (9) and MolProbity (10). Asn139 is the only residue in the disallowed region of the Ramachandran plot. For details of the data collection and refinement statistics, see Table S1.

Analytical ultracentrifugation. Sedimentation velocity experiments were performed at 4 °C in a Beckman Optima XL-I analytical ultracentrifuge. Double-sector cells were loaded with the protein sample of Rae1•Nup98^{GLEBS} (in a solution containing 50 mM Tris, pH 8.0, 0.15 M NaCl, 2 mM TCEP) and the reference solution (50 mM Tris, pH 8.0, 150 mM NaCl, and 2 mM TCEP), respectively. Data were recorded with absorbance detection at a wavelength of 280 nm. The partial specific volume and the solvent density were calculated using the SEDNTERP program. The SEDFIT analysis program was used to analyze the absorbance profiles and to calculate the sedimentation coefficient distribution, c(s), which was then transformed into a molar mass distribution, c(M) (11).

Protein binding experiments. Purified GST-Nup98^{GLEBS} wild-type and mutant proteins were loaded onto glutathione resin in a buffer containing 10 mM Tris, pH 8.0, 150 mM NaCl, and 0.5 mM TCEP. Purified Rae1 was added, and binding was allowed to proceed for 30 min at 4 °C. Beads were washed extensively with the same buffer, and bound proteins were analyzed using Coomassie-stained SDS-PAGE gels.

Yeast strains. A haploid strain containing Gle2-GFP at its chromosomal location was purchased from the yeast-GFP clone collection (Invitrogen). The Nup116 open reading frame was replaced with the kanMX6 cassette by homologous recombination in the presence of the plasmid pRS416-mCherry-Nup116, which was subsequently shuffled out by 5-fluoroorotic acid (5-FOA) (Zymo Research) after the pRS415-mCherry or the pRS415-mCherry-Nup116 (wild-type, ΔGLEBS, or E154K/E155K) plasmid was introduced. The transformants were selected twice on SD-LEU plates containing 5-FOA to ensure the loss of the full-length pRS416-mCherry-Nup116 plasmid prior to the analysis of various Nup116 constructs. The Nup116 deletion was introduced into the wild-type strain, BY4741, as well. The resulting *Δnup116* strain was used for transformation with the plasmid pRS415-mCherry or pRS415-mCherry-Nup116 (wild type or mutants). The details of the yeast expression constructs are listed in Table S2.

Yeast growth assay. The yeast strains carrying Gle2-GFP and mCherry-Nup116 variants were grown to midlog phase in SD-LEU media and diluted to 10 million cells/mL. This stock was used to generate a 10-fold dilution series, of which 10 μL were spotted on SD-LEU plates and grown at 23, 30, and 37 °C for 2–3 days.

In vivo localization. The yeast strains carrying Gle2-GFP and mCherry-Nup116 variants were grown in selective medium at 30 °C, and the live cells were analyzed with fluorescence microscopy, using a Carl Zeiss AxioImagerZ.1 equipped with an AxioCamMRm camera.

Fluorescence in situ hybridization (FISH) assay. The FISH experiments were carried out as previously described (12). Briefly, liquid cultures of yeast strains carrying mCherry-Nup116 variants were grown at 30 °C in SD-LEU media to an OD_{600} of 0.4. The cultures were shifted to 37 °C for 3 h prior to fixation in formaldehyde. The cells were analyzed by FISH using an Alexa-647-labeled 50-mer oligo dT probe. The statistical analysis was carried out using four independent images per strain.

Electrophoretic mobility shift assay. Two micromolar of a degenerate decameric RNA oligonucleotide was incubated with increasing concentrations of the Rae1•Nup98^{GLEBS} complex in a buffer containing 10 mM Tris, pH 8.0, 150 mM NaCl, 0.5 mM TCEP, and 5% glycerol at room temperature for 5 min. Samples were separated on a 6% native PAGE gel that was prepared with 45 mM Tris, pH 8.5 (titrated with boric acid to allow Rae1•Nup98^{GLEBS} to enter the gel) and prerun in the same buffer. After electrophoresis, the RNA was visualized through the use of SYBR Gold nucleic acid gel stain (Invitrogen).

Illustration and figures. Figures were generated using PyMOL (13). The electrostatic potential was calculated using the Adaptive Poisson–Boltzmann Solver (14). Sequence alignments were generated using ClustalX (15) with some adjustments and colored with Alscript (16).

1. Otwinowski Z, Minor W (1998) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* 276:307–326.
2. Collaborative Computational Project Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D* 50:760–763.
3. Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* 64:112–122.
4. Bricogne G, Vonrhein C, Flensburg C, Schiltz M, Paciorek W (2003) Generation, representation and flow of phase information in structure determination: Recent developments in and around SHARP 2.0. *Acta Crystallogr* 59:2023–2030.
5. Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 47:110–119.
6. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126–2132.
7. Brunger AT, et al. (1998) Crystallography & NMR System (CNS): A new software system for macromolecular structure determination. *Acta Crystallogr D* 54:905–921.
8. Murshudov GN (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* 53:240–255.
9. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291.
10. Davis IW, et al. (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35:W375–383.
11. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* 78:1606–1619.
12. Seo HS, et al. (2009) Structural and functional analysis of Nup120 suggests ring formation of the Nup84 complex. *Proc Natl Acad Sci USA* 106:14281–14286.
13. DeLano WL (2002) The PyMOL Molecular Graphics System (DeLano Scientific, San Carlos, CA) www.pymol.org.
14. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98:10037–10041.
15. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998) Multiple sequence alignment with Clustal X. *Trends Biochem Sci* 23:403–405.
16. Barton GJ (1993) ALSCRIPT: A tool to format multiple sequence alignments. *Protein Eng* 6:37–40.
17. Belgareh N et al. (2001) An evolutionarily conserved NPC subcomplex, which redistributes in part to kinetochores in mammalian cells. *J Cell Biol* 154:1147–1160.
18. Hsia KC, Stavropoulos P, Blobel G, Hoelz A (2007) Architecture of a coat for the nuclear pore membrane. *Cell* 131:1313–1326.
19. Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA* 89:10915–10919.

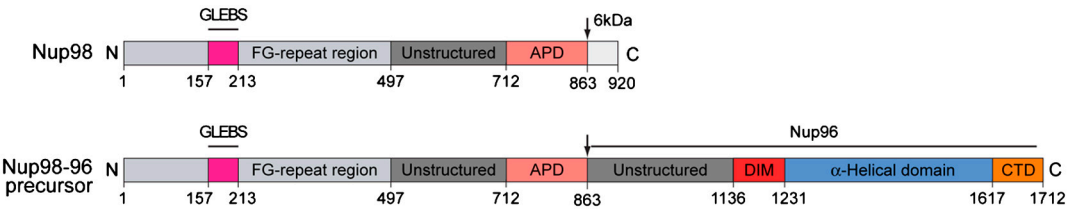

Fig. S1. Domain structure of Nup98 and the Nup98-96 precursor. For Nup98, the GLEBS motif (magenta), the phenylalanine-glycine (FG)-repeat region (gray), the unstructured region (dark gray), the autoproteolytic domain (APD; pink), and the C-terminal 6 kDa fragment (light gray) that is removed by cotranslational proteolysis are indicated. In the Nup98-96 precursor, an alternatively spliced version of Nup98, the 6 kDa fragment is replaced by Nup96, a protein that is embedded in the symmetric core of the nuclear pore complex (17, 18). The arrows indicate the sites of autoproteolytic cleavage.

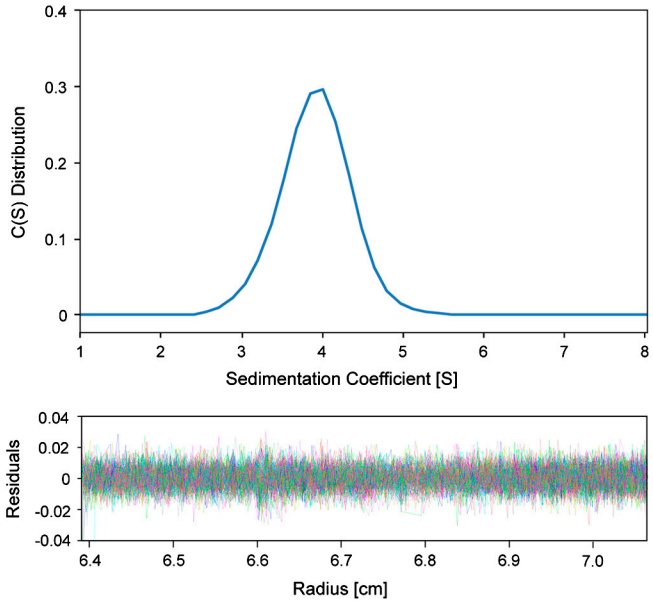

Fig. S2. Sedimentation velocity analysis of the Rae1•Nup98^{GLEBS} complex. The sedimentation coefficient distribution shows a single species (*Top*). The determined molecular mass is 45.4 ± 7.5 kDa, corresponding to a monomeric species of the Rae1•Nup98^{GLEBS} heterodimer (calculated 48.4 kDa) in solution. (*Bottom*) Residuals for the fitting.

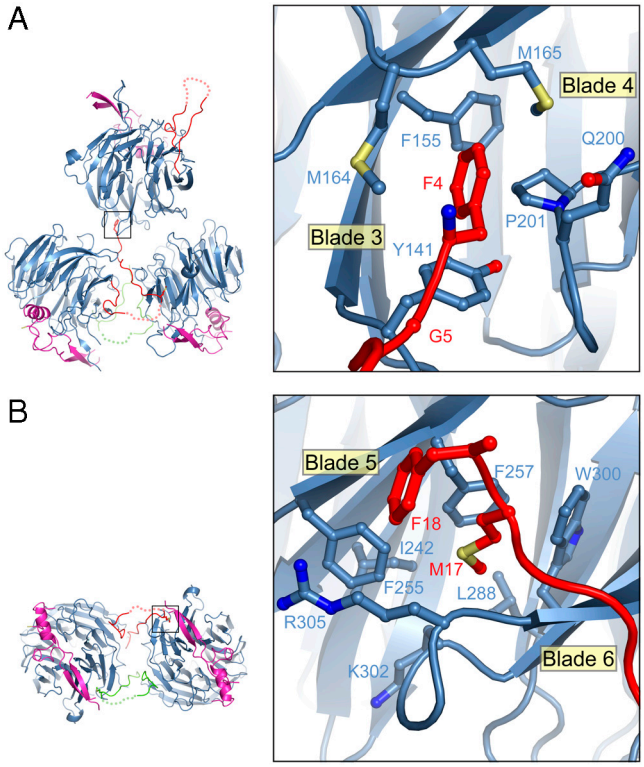

Fig. S3. Crystallographic interactions mediated by the N-terminal extended peptide segment (NTE) of Rae1. (*A*) Interactions between the four Rae1•Nup98^{GLEBS} complexes in the asymmetric unit are illustrated. For clarity, the fourth copy of the complex is omitted. The N-terminal FG-repeat-like sequence (residues 4 and 5) of a Rae1 molecule binds to a hydrophobic pocket of an adjacent Rae1 molecule. A close-up view of the FG-binding pocket is shown in the right inset. (*B*) Two Rae1•Nup98^{GLEBS} complexes interact via a handshake mode mediated by the Rae1 NTE. A 90° rotated view of the bottom two molecules of *A* is shown. The MFG sequence stretch (residues 17–19) of the Rae1 NTE binds to a hydrophobic pocket of a neighboring Rae1 molecule. A close-up view of the interaction is shown in the right inset. No electron density is observed for Gly19 of Rae1.

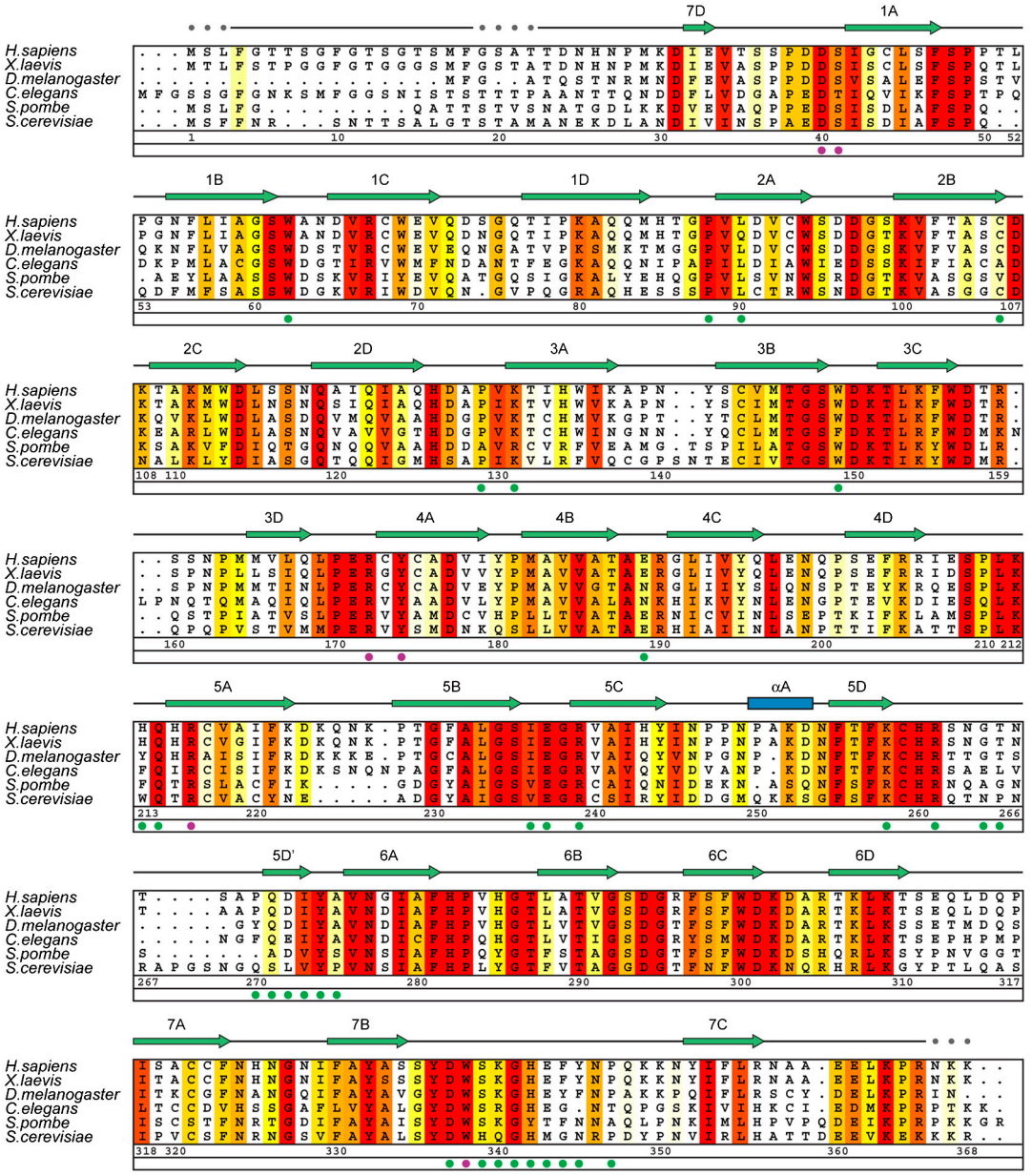

Fig. S4. Multispecies sequence alignment of Rae1 homologs. The numbering of the residues and the secondary structure are according to human Rae1. The secondary structure is indicated above the sequence as green arrows (β-strands), blue rectangles (α-helices), gray lines (coil regions), and gray dots (disordered residues). The overall sequence conservation at each position is shaded in a color gradient from yellow (60% similarity) to red (100% identity) using the Blosum62 weighting algorithm (19). The Rae1 residues that are involved in the interaction with Nup98 GLEBS motif are indicated with dots below the alignment. The residues that are labeled with magenta dots are shown as sticks in Fig. 4*B*.

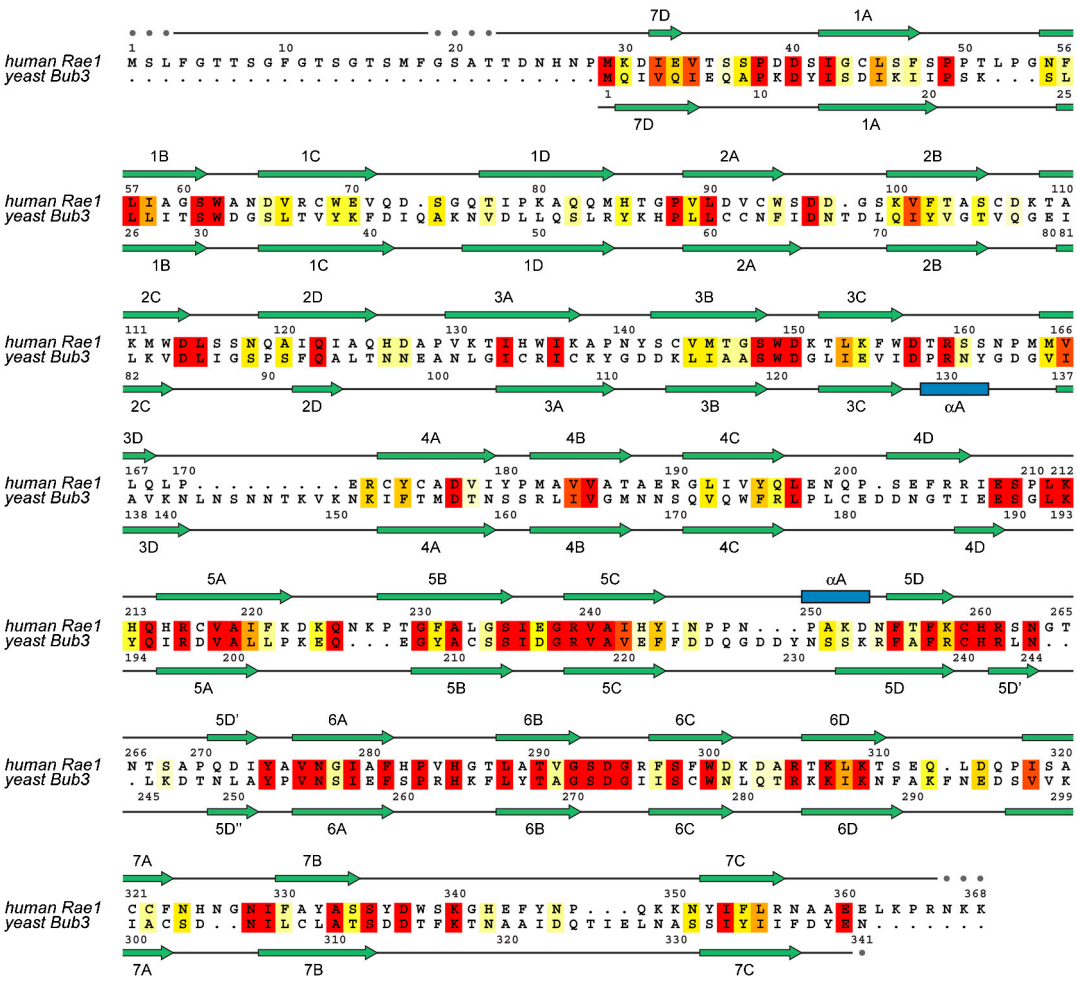

Fig. S5. Structure-based alignment of the primary sequence of human Rae1 and yeast Bub3. Corresponding secondary structure elements of the two β-propellers are indicated above (Rae1) and below (Bub3) the aligned sequences.

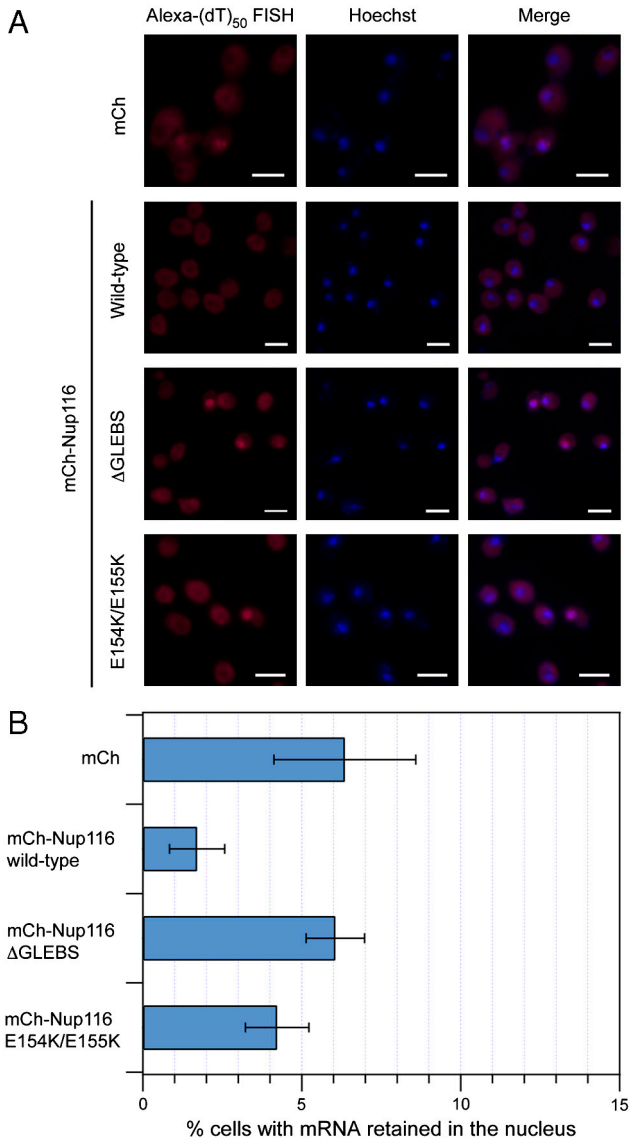

Fig. S6. mRNA export assay of Nup116 mutants. (*A*) Detection of poly(A) mRNA and DNA was carried out using an Alexa-647- labeled 50-mer oligo dT FISH probe and the Hoechst DNA stain, respectively. The merging of the two images is shown on the right. (Scale bar: 5 μm.) (*B*) Quantitation of nuclear poly(A) mRNA retention. The percentages refer to the fraction of cells displaying marked nuclear staining. The error bars correspond to the standard deviation and are derived from four independent images. The P values of the Nup116 ΔGLEBS and E154K/E155K mutants are less than 0.05, compared to the wild-type Nup116 construct (0.00048 and 0.00936, respectively), indicating that the observed differences between these strains is statistically significant.

Table S1. Crystallographic analysis.

	Crystal 1 native	Crystal 2 OsO_4
Data collection		
Synchrotron	APS	APS
Beamline	GM/CA-CAT 23ID-B	GM/CA-CAT 23ID-B
Space group	P1	P1
Cell dimensions		
a, *b*, *c*, Å	$a = 56.4$, $b = 79.3$, $c = 93.4$	$a = 56.4$, $b = 79.1$, $c = 93.8$
α, β, γ, °	$\alpha = 76.6$, $\beta = 90.0$, $\gamma = 89.9$	$\alpha = 76.2$, $\beta = 89.8$, $\gamma = 89.7$
Wavelength, Å	1.14014	1.14013
Resolution, Å	20.0–1.65 (1.71–1.65)	20.0–3.0 (3.11–3.0)
R_{sym}, %	7.8 (65.6)	18.4 (30.5)
$\langle I/\sigma I \rangle$	15.5 (1.8)	7.3 (3.0)
Completeness, %	95.6 (87.1)	96.7 (83.5)
Redundancy	3.8 (3.1)	3.4 (2.3)
Refinement		
Resolution, Å	20.0–1.65	
No. reflections		
Total	170,317	
Test set	9,058 (5.0%)	
R_{work}/R_{free}, %	20.6/23.7	
No. atoms	13,611	
Protein	12,765	
Water	798	
Ligand	48	
B factors	27.5	
Protein	27.1	
Water	31.3	
Ligand	96.7	
rms deviations		
Bond lengths, Å	0.010	
Bond angles, °	1.26	
Ramachandran statistics		
Most favored, %	88.1	
Additionally allowed, %	11.5	
Generously allowed, %	0.1	
Disallowed, %	0.3	

Highest-resolution shell is shown in parentheses

Table S2. Expression constructs

Baculoviral expression constructs

Protein	Residues	Expressionvector	Restriction sites 5′, 3′	N-/C-terminal sequence overhang
Rae1	1–368	pFastbac Dual	SmaI, KpnI	None/HHHHHHH
Nup98 GLEBS	157–213		BamHI, EcoRI	GS/None
Rae1	1–368	pFastbac Dual	SmaI, KpnI	None/HHHHHHH

Bacterial expression constructs

Protein	Residues	Expression vector	Restriction sites 5′, 3′	N-terminal sequence overhang
Nup98 GLEBS	157–213	pGEX-4T1	BamHI, EcoRI	GST-LVPRGS
Nup98 GLEBS-N	157–179	pGEX-4T1	BamHI, EcoRI	GST-LVPRGSGGS
Nup98 GLEBS-C	179–213	pGEX-4T1	BamHI, EcoRI	GST-LVPRGSGG
Nup98 GLEBS E201K, E202K	157–213	pGEX-4T1	BamHI, EcoRI	GST-LVPRGS

Yeast expression constructs

Protein	Residues	Shuffle vector	Promoter	Restriction sites 5′, 3′
Nup116	1–1113	pRS415-mCh	Nop1	NotI, SacII
Nup116 ΔGLEBS	1–1113 residues removed 115–166	pRS415-mCh	Nop1	NotI, SacII
Nup116 E154K, E155K	1–1113	pRS415-mCh	Nop1	NotI, SacII